T A M P E R E   E C O N O M I C   W O R K I N G   P A P E R S

N E T   S E R I E S

# LINKED EMPLOYER-EMPLOYEE DATA ON FINNISH PLANTS FOR THE ANALYSIS OF PRODUCTIVITY, WAGES AND TURNOVER

Pekka Ilmakunnas
Mika Maliranta
Jari Vainiomäki

Working Paper 1
April 2001
http://tampub.uta.fi/econet/wp1-2001.pdf

DEPARTMENT OF ECONOMICS

FIN-33014 UNIVERSITY OF TAMPERE, FINLAND

ISSN 1458-1191
ISBN 951-44-5091-4

# LINKED EMPLOYER-EMPLOYEE DATA ON FINNISH PLANTS FOR THE ANALYSIS OF PRODUCTIVITY, WAGES AND TURNOVER

Pekka Ilmakunnas
Helsinki School of Economics

Mika Maliranta
The Research Institute of the Finnish Economy (ETLA) and Statistics Finland

Jari Vainiomäki
University of Tampere

Tampere Economic Working Papers
Net Series
Working Paper 1
April 2001
http://tampub.uta.fi/econet/wp1-2001.pdf

ACKNOWLEDGEMENTS

This research has been funded by the Finnish Ministry of Labour and Academy of Finland. The work for this paper has been done following the terms of data usage authorization by Statistics Finland. We thank Statistics Finland personnel for their help in data usage. This is a revised version of a paper prepared for Nordic Workshop on Labour Market Research with Register Data, 10.-11.4.2000 Tampere, Finland.

# LINKED EMPLOYER-EMPLOYEE DATA ON FINNISH PLANTS FOR THE ANALYSIS OF PRODUCTIVITY, WAGES AND TURNOVER


## Abstract

The process of linking various registers for the analysis of Finnish plants is described and potential uses of the linked data are illustrated. Information on individuals from the Employment Statistics is matched with information on plants from the Business Register and the Industrial Statistics on the basis of plant codes. Also firm-level information from the R&D statistics and the Financial Statements Statistics is linked to the plants. Various steps in the linking are described and the quality of the matched data is evaluated. Somewhat different linking is used for creating data sets for different research purposes. One data set is a panel of manufacturing plants. Information on the average education, age, and seniority of the employees and on their turnover is linked to data on production and input use. This data can be used for the analysis of the impact of employee characteristics on productivity and for the comparison of wage and productivity profiles by age and experience. Other linked data sets have been formed for both manufacturing and service sectors plants to examine worker and job turnover simultaneously and their dependence on plant and employee characteristics.



Pekka Ilmakunnas
Department of Economics
Helsinki School of Economics
P.O. Box 1210, FIN-00101 Helsinki, Finland
ilmakunn@hkkk.fi

Mika Maliranta
The Research Institute of the Finnish Economy (ETLA)
Lönnrotinkatu 4 B, FIN-00120 Helsinki, Finland
Mika.Maliranta@etla.fi

Jari Vainiomäki
Department of Economics
FIN-33014 University of Tampere, Finland
Jari.Vainiomaki@uta.fi

# 1. Introduction

The creation and use of matched employer-employee data is essential for the analysis of many industrial and labor market issues (see e.g. Haltiwanger et. al., 1999). However, access to this type of information has been rare. In the present paper we discuss the linking of various registers on firms, plants and employees to analyze the Finnish labor market. We describe the various data sources used in the linking of employees and employers, the process of linking, and the problems encountered. In this work, the solutions made in the production process of the official statistics restrict to some extent the freedom of the researchers who use the data. We discuss how different practices in the various statistics lead to incompleteness of the matching. The research topics themselves may lead to a further loss of data if some key variables are available in sufficient accuracy only for a subset of plants or individuals.

As an illustration of the uses of this kind of data sets, we briefly describe two larger research projects where the linked data has successfully been used. The first one examines the productivity of manufacturing plants. The interesting issues are how one can explain heterogeneity in plant productivity with the fact that the 'quality' of the work force varies across plants, and whether wage formation is based on productivity or on incentive considerations. To study these questions, we need to link information on employees to plant data. A major problem in the research on these questions has been the difficulty of measuring the productivity of individuals, although their earnings can be measured with reasonable precision. However, register-based data sets that match information on individual employees and their employers provide a way of measuring the productivity profile of workers. We assess the effect of various human capital components on productivity by using panel data from the Finnish manufacturing plants that is extended with variables measuring average employee characteristics as well as plant-level measures of worker turnover and other plant characteristics.

The second illustration deals with job and worker flows. Analysis of job creation and destruction using longitudinal data on plants or firms has become increasingly popular in labor market and macroeconomic research. However, only seldom can the simultaneous inflow of workers to the plants and outflow of workers from the plants be calculated in a similar way to form consistent measures of excessive worker turnover or churning and to study their dependence on plant and employee characteristics. Some researchers have used some specialized firm surveys, unemployment insurance records, etc. However, in Finland, like in the other Nordic countries, there are complete registers on individuals and firms, which can be linked for this kind of analysis. The topic is highly relevant in light of the very deep recession in Finland in the 1990s, which provides a good opportunity to examine the cyclical properties of the flows.

Most of the paper is devoted to discussion on the process of linking various registers and the quality of the resulting data sets. Results on the topics described above are presented as an illustration on the research potential of the data sets. More detailed results are presented in separate reports.

In Section 2 of this paper we describe the main data sources and the main principles of the linking process. In Section 3 we examine the properties of the linked data sets. In Section 4 we present some results, and Section 5 concludes the paper.

## 2. Data Sources and the Linking Process

### 2.1 An Overview of the Linking Process

The unique identification codes for persons, enterprises and plants used in different registers form the backbone of the Finnish register network whereby different sources of information can be integrated conveniently for statistics purposes. The process of linking the registers is illustrated in Figure 1. Business Register (BR), Employment Statistics (ES) and Industrial Statistics (IS) are three basic registers maintained by Statistics Finland that are relevant for the current exercise to build linked employer-employee data sets. These are the primary data sources for business units, worker characteristics, and industrial plant characteristics, respectively. The arrows between these boxes denote connections through plant codes. In the next stage (step 4) the ES and BR data are linked to two plant-level data sets. PESA (Plant-level Employment Statistics Data on Average Characteristics) is a data set that contains information on the average characteristics of the employees in each plant and PESF (Plant-level Employment Statistics Data on Flows) contains information on the employment change (job flow) and inflow and outflow of workers in each plant. These data sets include all non-farm (and non-forestry) business sector plants. Our definition of the non-farm business sector consists of minig (C), manufacturing (D), energy etc. (E), construction (F), trade (G), hotels and restaurants (H), transportation etc. (I), finance (J), and real estate, business services etc. (K). Hence, agriculture, forestry and fishing (A,B), public administration (L), education (M), health and social work (N), other social and personal services (O), international organizations (Q), and industry unknown (X) are excluded.

For industrial (i.e., manufacturing, mining and energy) plants, the IS data are used for calculating various plant-level characteristics, which gives data set LDPM (Longitudinal Data on Plants in Manufacturing). Finally (step 9), the data sets PESA and PESF can be linked to some firm-level information from the R&D Statistics (RDS) and Financial Statements Statistics (FSS). In the case of industrial plants, also the LDPM data are linked to the other data. The details of the registers, the linking process, and the data sets are explained below.

**FIGURE 1 HERE**

## 2.2. The Registers

The **Employment Statistics** database compiles information on the economic activity of individuals and their background characteristics from a large number of different administrative registers. It covers effectively the whole population of Finland. There are over 2 million employees in this register. In the business sector (defined above in footnote 1) there are more than 1.1 million employees in about 100 000 plants. The enterprise and plant identification codes, industry and other general information needed in Employment Statistics are taken as such from Business Register. The employer-employee links on which our linked data rest are those determined in the Employment Statistics system. The employer-employee match in Employment Statistics is based on the Register of Wages and Pensions, which includes information on all employment spells during a year of all individuals in Finland and is a part of the Employment Statistics production system. For each person a unique plant appearing in Business Register is determined as his/her primary employer during the last week of each year. This connection is traced out using the enterprise identification codes in the Register of Wages and Pensions. For multi-unit enterprises the person-plant matches are determined using a questionnaire asking enterprises to attach persons to their appropriate plants. Furthermore, checkups and corrections are performed by comparing the geographical location of plants and the place of residence of persons. Linking an individual with the proper employer plant is a challenging task, and there remain a number of persons in Employment Statistics whose plant code is missing or may be improper. However, a great deal of effort is being made in Employment Statistics to seek the correct plant linkage for each individual, so we consider this information to be the best available on which a linked employer-employee data can be built. The Employment Statistics database was started in 1987, but on the basis of preliminary investigations and discussions with the Employment Statistics department, there were suspicions about the data quality in the first year of the database.

One could use the individuals as basic units and link the plant characteristics to them in order to study, for example, the determinants of individual incomes and changes in employment status. Alternatively, the Employment Statistics can be used as a source of information on the characteristics of each plant's work force. We have formed a plant-level panel data set PESA from the information on individuals by calculating plant-level sums or averages of the background characteristics (age, education, seniority, etc.) of the employees. Another plant-level data set, PESF, includes information on employment, and inflow and outflow of workers. These data sets are then linked to plant-level data from the other registers. The main features of ES and the other registers are summarized in Table 1.

**TABLE 1 HERE**

The **Business Register** data base of Statistics Finland covers registered employers and enterprises subject to VAT and their plants in Finland, and it is the basic source of enterprise and plant codes used in other Statistics Finland registers and statistics. In 1998 there were over 200000 business sector plants in the register. Identification codes for enterprises used in Business Register originate from tax authorities. Identification codes for plants, in turn, are given by Business Register when a new plant is established. Business Register also follows changes in the demographic structure of plants and enterprises like their death and changes in ownership. Furthermore, Business Register includes information on the contact address, classifications like industry, and some basic variables like sales, employment, and the wage bill. In the analysis of job and worker flows, this is the source of information on wages and industry classification. However, the information content of Business Register is fairly limited, so that for the analysis of industrial plants it is supplemented from other sources.

The **Industrial Statistics** compiles comprehensive information on the economic activity of industrial plants by annual surveys. When a plant in Business Register fulfills the selection criteria to be included in the Industrial Statistics Survey (employing at least five person being the main criterion up to the year 1994), it is picked into the information system of Industrial Statistics. The enterprise and plant identification codes, industry group, etc. originate at this stage from Business Register. However, since then the plant's identification codes, classifications and contact information are maintained and updated, if need be, in the systems of Industrial Statistics. Therefore, it is possible that the connection with the plant's original counterpart in Business Register may weaken or disappear over time, which causes some problems when matching Industrial Statistics with other data sources that use the codes from Business Register (see below for more details). Industrial Statistics is our main source for plant-level variables, like output, capital stock, and hours worked. The employment figures in Industrial Statistics represent average employment during the year. The plant-level data series from Industrial Statistics are available for the period 1975-1994. After 1994 there is a major break in the data collection practices, which dictates the final year. The full Industrial Statistics data base includes annually about 7 000 - 8 000 plants, but in our analysis we concentrate on active production plants (omitting, e.g., headquarters and auxiliary units), so our basic plant data set, LDPM, includes approximately 6 000 plants annually.

The **R&D Statistics** contains information on the R&D expenditure of firms. As far as the past few years are concerned, this data source covers all large firms (employing at least 100 persons). In addition, questionnaires are sent to a group of 'potential' R&D firms. This information is obtained from various registers related to R&D activities that are available to Statistics Finland. Finally, the data set on R&D activities of firms is extended by a random sample of smaller firms. Before the year 1995 R&D survey was based on the same sample of firms as Financial Statements Statistics.

Nowadays the R&D survey data set includes nearly 3 000 firms (or, in some cases, groups of enterprises). Most of the firms are from manufacturing (1 500 – 2 000 firms), but there are also some service sector firms. The R&D data are used in the analysis of job and worker flows as explanatory variables. The firm codes in R&D Statistics are the same as in Business Register. The information on the firm's R&D expenditure can therefore be linked to those plants that belong to the firm. However, this can be done only for a subset of all plants.

The **Financial Statements Statistics** has information on the balance sheets and income statements of firms. The large and medium-sized firms, basically those employing at least 20 persons, are covered each year. In the past few years, information on smaller firms has been obtained from registers compiled for taxation purposes. In earlier years a rotating sample of the smaller firms was applied. Each year, some 10 000 firms are included in the survey. This data are used as background information in the analysis of job and worker flows. Again, the firm codes are the same as in Business Register, which can be used for linking data on profitability of the parent firm to the plants.

## 2.3 Plant and Firm Codes

When all the non-farm business sector plants are examined, the worker characteristics can be linked to the plants by using the plant codes in Employment Statistics and Business Register. In principle, these codes are from the same source. However, there exist some codes in both sources that cannot be found in the other source. There are many plants in Business Register that cannot be found in Employment Statistics. Usually these are extremely small so that the linked data set covers a large share of the total employment in Business Register. There are also fairly many plant codes in Employment Statistics that do not appear in Business Register. These are typically some kind of 'auxiliary' codes created in the cases where actual employer-employee link cannot be traced. Outcomes of the linking process are described in a greater detail in Table 5 in Section 3.2. Sometimes it is useful to link firm-level data to plant-level data. In these cases the linking is performed by using a firm code that appears for example in Financial Statements Statistics data or R&D Statistics data. Information on this sort of linking is provided in Tables 6 and 7 in Section 3.2.

When the data on worker characteristics in industrial plants are matched with plant data from Industrial Statistics, more difficulties are encountered because of differences in the plant codes in the two sources. They originate from the same source, but after the initial appearance of a new plant in Business Register, the plant codes are maintained in the Industrial Statistics system. Since the objectives and data needs differ between Business Register and Industrial Statistics this may lead to some differences in plant delineation and plant identification codes in these two systems. Industrial Statistics strives for providing a comprehensive description of the industrial activities in different industries and regions. To this end the information should ideally be surveyed from a unit that engages in one or predominantly one kind of activity at a single location, which is the basic 'definition' of a plant in Industrial Statistics. Although in theory an establishment-based survey is conducted for Industrial Statistics, in practice an 'establishment-type of unit' may be used, which means in many cases some kind of mixture of a local unit and a kind-of-activity unit. In some special cases it is even allowed that an integrated whole that is defined as the statistical unit, consists of parts locating geographically detached. When the unit covers such an integrated whole it is often easier for an enterprise to give comprehensive, relevant and reliable information for the purposes of Industrial Statistics. Business Register in turn keeps record of the 'population' of business units with a limited information content on enterprises and establishment, so it is substantially easier for Business Register to stick to a stricter definition of an establishment unit.

There are also some differences in treating demographic events in Business Register and Industrial Statistics. Generally, Industrial Statistics is more reluctant to change the code of a unit that is continuing activities after a demographic event. For example, if two or more units are merged, generally the code of the oldest plant is kept and the other are incorporated under it in Industrial Statistics. However, this treatment is not fully formal and also the size and the industry of the units are considered in making the decision. On the other hand, when the Business Register considers changing a plant code, it takes into account three criteria: industry, address and ownership. In principle, the plant code is changed if at least two of the above criteria change. However, in practice these criteria have been used only as guidelines for decisions made case by case. Plant code may be renewed, if a 'substantial' change has occurred only in industry or only in location. When a plant is transferred to a new owner (simple change in ownership), the plant code does not change. In cases where two or more plants are combined, the practices in Business Register have varied to some extent. In some cases, a new plant code is given to the new combined unit.

Because Industrial Statistics generally follows a more conservative policy than Business Register in changing the plant codes, there are a number of codes in Industrial Statistics that cannot be found in Business Register. Especially the older plants (which are likely to be larger as well as likely to have been involved in demographic changes) are exposed to a greater risk that the connection with any Business Register code, and therefore with Employment Statistics, is broken down. Conversely, there are a large number of plant codes in Business Register and Employment Statistics that cannot be found in Industrial Statistics, because they do not fall within the criteria defined for the units to be included in the Industrial Statistics survey (i.e., the small plants employing less than five persons). Furthermore, due to differences in definitions of plant delineation, a plant in Industrial Statistics may have a specific Business Register code, but in practice is a composite of several Business Register plants. In principle, also a converse situation is possible, where a Business Register plant is divided into two separate plants in Industrial Statistics. These differences in plant coding practices clearly cause some matching problems when using plant codes from Business Register and Industrial Statistics. However, the differences should not be given too much emphasis. In most cases, there are no discrepancies between the two systems, and the simple reliance on ‘administrative’ plant codes yields a correct match. More refined procedures to unify plant coding in Business Register and Industrial Statistics, based on historical records of coding changes and/or using data on individuals to form consistent plant identifiers, is a major task and was considered outside the present work.

## 2.4. Construction of Variables on Plant and Worker Characteristics

The list of variables available in the register-based Employment Statistics is too extensive to go through in full length here. Among other things, for each person the following information is included or can be inferred: personal identification code, identification code for the employer enterprise and plant, industry of the plant, age, marital status, education (Statistics Finland educational classification in great detail), experience (general and firm-specific), income from employment, other income, and labor force status (employed, unemployed or out of the labor force). In other words, this data set offers many opportunities for investigating interesting hypotheses about the connection of employee characteristics and plant performance. In this paper, we take interest especially in such characteristics as age, education, experience, and changes in employment or labor force status. For those plants from which we have information on at least two employees, we have calculated the following average employee characteristics (in years): age, experience in the plant, and schooling. The schooling years are based on detailed information on the educational degrees held by persons, which are transformed to years using years to complete the degree. The worker characteristics data set for industrial plants is the same as that used in Vainiomäki (1999), but aggregated to plant totals for the present work.

We have also measures of worker flows for each plant during successive pairs of years from the Employment Statistics database. We have calculated the number of persons who appear in the same plant in both years (stayers). Similarly, we have counted those who have exited (worker outflow or separation) and those who have entered during the period (worker inflow or hiring). The difference of inflow and outflow is the net change of employment. The inflows and outflows were calculated each year for each plant. Further, plants were classified into industries or other groups, determined by different background variables. In each industry or group, the worker inflow rate or hiring rate (WIF) and the worker outflow rate or separation rate (WOF) were calculated by dividing the respective flows by the average employment in two successive years in the industry or group. (This scaling follows the suggestion of Davis, Haltiwanger and Schuh, 1996.) The worker flow rate or worker turnover rate (WF) is the sum of WIF and WOF, and the net rate of employment change (NET) is the difference of WIF and WOF. The worker flows can also be calculated by source or destination. In particular, the worker inflow rate from unemployment (WIFU), the outflow rate to unemployment (WOFU), and the corresponding net flow from unemployment, UNET = WIFU – WOFU, can be examined.

Correspondingly, in each industry or group of plants the sum of positive net employment changes is job creation and the sum of the absolute values of negative employment changes is job destruction. When we divide them by average employment in two successive years, we obtain the job creation rate (JC) and job destruction rate (JD), respectively. Their sum is the job reallocation rate or job turnover rate (JR), and the difference of the job reallocation rate and absolute value of net change is the excess job reallocation rate, EJR = JR - |NET|. Finally, the churning flow rate (CF) measures excessive worker turnover. It is defined as the difference of the worker flow rate and job reallocation rate (Burgess, Lane, and Stevens, 2000), CF= WF – JR.

Business Register was used for calculating average wages, the wage bill divided by the number of employees, of the non-farm business sector plants for the analysis of job and worker flows. From Financial Statements Statistics we used only the profitability (net profits per sales, average of periods *t-1* and *t*) and from R&D Statistics R&D intensity (R&D expenditures per sales). Both were used as background variables in the analysis of job and worker flows.

Industrial Statistics includes a wide variety of variables on output and inputs of industrial plants. Output can be measured with gross output and value added. These variables were converted into real terms by using corresponding (2- or 3-digit) industry level implicit price indices obtained from the Finnish National Accounts. Labor and capital inputs are of particular interest in productivity analyses. The former can be measured by hours worked or the number of persons (separate figures for production workers and salaried staff are available). Since the number of employees includes for example temporarily laid off and those on maternity leave, it is an imperfect indicator of the labor input in production. Therefore, we used the actual hours worked as the labor input measure. As for capital stock measures, they have not been included in the questionnaires since 1985. Capital input estimates were derived for a vast majority of plants with a perpetual inventory method that makes use of investment figures in Industrial Statistics. Investments were converted into real terms with implicit price deflators obtained from National Accounts. Two estimates were constructed: one for machinery and equipment and another for buildings and constructions (see Maliranta, 1997, for details). It seems that the quality of the machinery and equipment measure is superior to that of buildings and constructions. As the capital services from the former are substantial, machinery and equipment capital is preferred as a proxy of the total capital input.

As we are seeking factors that affect the productive performance of the plants, we need a suitable indicator for it. The total factor productivity is a useful measure as it incorporates efficiency both in labor and capital usage. We measured total factor productivity directly using the translog multilateral productivity index introduced by Caves, Christensen, and Trethway (1981) and Caves, Christensen, and Diewert (1982). It allows the factor elasticities to vary across plants and industries. When using this procedure we are able to pool conveniently different industries (see details in Maliranta, 1997 and 1999).

The wage level of the industrial plants is the average wage in the plant calculated by dividing total wages paid by hours of employees. Other plant-level variables from Industrial Statistics include geographical location, the ratio of rents paid to the value of machinery, an indicator of foreign ownership, recent investments, an indicator for plants that are going to disappear ('the shadow of the death' à la Griliches and Regev, 1995), average hours per worker, and capacity utilization. For the analysis, we also classified the plants to groups according to their age. We formed six generation groups (cohorts) separately for each 4-digit industry on the basis of the order of appearance of plants to Industrial Statistics. The newest two groups are decile classes and the rest of the groups are quintile groups. The generation is indicated by dummy variables GENA (newest) to GENF (oldest).

## 3. Description of the Matching Process

### 3.1 Linked Data for the Analysis of Plant Productivity

The process of matching workers in Employment Statistics to plants in Industrial Statistics proceeded as follows. First, those persons in the full Employment Statistics data base were chosen who are over 15 years old, whose employer plant's industry is manufacturing, and for whom the plant identification code exists. This can be treated as the full Employment Statistics based 'population' of all manufacturing workers. The number of these employees has a downward trend, which has been strengthened by the recession in the 1990s. Starting with 446 000 in 1988, the number of employees reached its minimum of 346 000 in 1993 (Table 2, line 1). For several reasons, a matching plant in Industrial Statistics cannot be found for all these employees. First, Industrial Statistics includes only plants employing at least five workers, whereas Employment Statistics also includes workers in smaller plants. Second, the group of plants for which workers are linked is restricted to those plants that have production activities in Industrial Statistics (omitting plants that are headquarters, auxiliary units, etc.). Finally, due to some differences in plant coding in the two statistics, as discussed above, we include only those plants (and their employees) which had exactly the same plant code in both systems. These restrictions decrease the number of individuals in the data set by 92 000 in 1988 (line 2) and somewhat less towards the end of the period. Depending on the year, this is a drop of 18 to 23 percent in the number of employees.

Before the calculation of the employee based plant-level variables, individuals with very short spells of employment (under 1 month) and wage income that was likely to be erroneous (average monthly wage outside certain minimum and maximum bounds) were dropped. These restrictions amount to a further loss of persons, which was about 70 000 in 1988 but clearly less in the other years (line 3). The remaining employees were used in forming the plant-level variables on work force characteristics of the plants in Industrial Statistics. These linked employees account for some 65 to 78 percent of the employment figures in Industrial Statistics (line 4). In the process of the linking, we also lose some plants in Industrial Statistics because no employees can be matched to them. The number of plants lost varies from 540 to 780 plants, or from 8 to 12 percent of the number of active production plants in Industrial Statistics. The remaining plants cover around 90 percent of active plants (lines 5 and 6).

**TABLE 2 HERE**

Because of the loss of employees and plants in the matching, there may be problems with the representativeness of the linked data. First, the loss of employees is proportionately larger than the loss of plants, since the share of plants with linked employees is about 90 percent, but the share of linked employees is 65 to 78 percent of employment in these plants (as measured in Industrial Statistics). Either the plants that are lost in the linking process are larger than average, or the loss of employees is greater in larger plants. We explored this by examining how the number of employees of the plants differed as measured in Industrial Statistics and in Employment Statistics. The correlation between these two measures in 1994 was slightly over 0.8, so it seems that on average the matching is reasonably accurate. However, some plants are much larger according to Industrial Statistics than according to the matched employees, indicating that a substantial proportion of these plants' workers is lost in the linking process. These problems are, however, more prevalent in medium-sized and small plants than in large plants. It therefore seems that large plants are under-represented in the linked data, but for the linked large plants the matching is fairly accurate. The difference in definition of plant delineation, leading to different plant codes in the two systems, may be more likely to occur in large plants. This could explain their greater loss.

In cases where the number of linked workers is smaller than in Industrial Statistics it is plausible to think that plant-fairly well level variables of employee characteristics are based on a ‘sample’ of all employees in the plant. A further complication is that we sometimes have more matched workers in a plant compared to its employment in Industrial Statistics. One possible explanation for this is the different concept of employment in Employment Statistics (end of year situation) and in Industrial Statistics (annual average employment). Second, employees have been linked to one plant only, although they may have a second job at another plant. Finally, it is possible that the attachment of persons to plants in the Employment Statistics data base is simply incorrect despite all efforts to form correct matching, or the difference in plant coding systems between the two sources causes some incorrect matching. Despite these problems, these linked data provide a rich source of information on the structure of the labor force of the plants, which is unparalleled to information from any other sources. It is also likely that the matching process reduces possible errors-in-variables problems, since one reason for incomplete matching may be data errors.

Some descriptive statistics on plant characteristics at various data steps are presented in Table 3 for the years 1990 and 1994. The sample of plants shrinks step by step as we are making more requirements for the content of the variables. The starting point is **data set A** that includes all plants in Industrial Statistics. **Data set B** excludes other plants than those having industrial activities (headquarters, auxiliary units etc.). This is the subset of plants in Industrial Statistics for which linked worker characteristics were searched. The share of active plants from all plants is about 75 to 85 percent, but in terms of employment they are larger than average, so their employment share is about 95 percent each year. **Data set C** retains only plants for which it was possible to construct a total factor productivity indicator. At this stage, the number of plants drops substantially (to about 60 to 70 percent) and average characteristics change for many reasons. First, an appropriate estimate of capital input (machinery stock), which is needed for the total factor productivity (TFP) measure, is lacking for a number of plants. These plants are typically smaller than average, so average plant size increases and the employment share remains at about 85 to 90 percent. Furthermore, we constructed the TFP indicator only for those plants whose ln(real value added per hour) and ln(real value added per machinery stock) did not differ too much from the corresponding industry average. If the value of either of these two indicators differs from the group average by more than 4.4 standard deviations, the plant is considered to be an outlier. Finally, outliers of the TFP index are picked out in a similar manner. Applying these restrictions leads to increased average labor productivity. **Data set D** is the subset of plants used in the regression analysis, and requires that appropriate estimates of both total factor productivity and characteristics of work force are available. The average size of plants increases somewhat but the other plant characteristics do not change. It is notable that labor productivity increases somewhat in each data step, except the last, but there are hardly any differences between the data sets in average wage per worker.

**TABLE 3 HERE**

Table 4 presents average worker characteristics for two sets of plants. First, we required at least two individuals to calculate the average labor characteristics for the plant. In 1994, there are 4 755 such plants with the average labor characteristics data, but only 3 882 of these plants were used in the regression analysis (data set D). The unavailability of an acceptable total factor productivity indicator (and other variables needed in the regressions) drops the number of plants considerably. Table 4 shows that average labor characteristics of plants change quite moderately due to this requirement. The average age (AGE) is practically the same in the sample of plants used in the regression analysis as among all linked plants. In addition, the distributions are similar. In both data sets there are some two hundred plants annually where the average age of personnel is less than 30 years or more than 45-46 years. The difference in seniority between the two samples of plants is somewhat more notable. There is a wide range of variation in the seniority variable (SEN) across plants, the 5th percentile being about two years and the 95th some 17- 19 years depending on the year. As for schooling years (EDUY), the difference between the samples is insignificant. The great majority of the plants fall in the range from 9.5 to 12 years of average schooling of employees.

**TABLE 4 HERE**

## 3.2 Linked Data for the Analysis of Job and Worker Flows

For the purpose of analyzing job and worker flows in different sectors and in different kinds of plants we have used various data sets. Employment Statistics is the most important source of information. What is crucial for our purposes is that it includes (at least for the main part of the individuals) firm and plant codes that identify the employer and workplace of the person and thereby facilitate the calculation of the flows. The plant-level flow data are linked to

Business Register plants by plant code. In order to have consistent job and worker flow series we have dropped those persons whom are not linked to a plant that appears in Business Register. On the other hand, we drop those Business Register plants, for which no employees can be found in Employment Statistics.

More extensive loss of plants happens when the plants are connected to Financial Statements Statistics and R&D Statistics. These statistics include only a subset of firms and hence also of plants. Therefore, when job and worker flows are calculated according to the profitability of the parent firm, or by the firm's R&D intensity, the aggregate flows are based on a sample of plants.

Further, in the case of industrial plants, some additional variables like capital intensity and export intensity were used as the basis of classifying plants into groups. This, first of all, restricts the examination to manufacturing. Secondly, within manufacturing, we restrict attention to those plants, for which e.g. a capital stock measure can be calculated.

We examine the representativeness of the resulting data sets by measuring the share of employment and plants that each of the data sets covers. First, Table 5 shows the outcome of linking Employment Statistics with Business Register at the plant level. According to Employment Statistics there have been 1.1 to 1.5 million employees in the business sector in the period from 1988 to 1996. The respective figure on the basis of Business Register on plants is 1.0 to 1.3 million employees. The difference in employment between the two sources can be explained at least partly by the fact that in Business Register the number is given in terms of fulltime workers whereas Employment Statistics uses the number of heads. Also the time period that the employment figures cover is different (end of year vs. annual average).

**TABLE 5 HERE**

We note from row 3 in Table 5 that our plant-level data on flows (PESF), derived from Employment Statistics linked with Business Register, accounts for about 80 percent of persons in Employment Statistics. In other words, there are annually more than 200 000 persons in the non-farm business sector that are not linked to a plant appearing in Business Register. There are also such plant codes that do not appear in Business Register (see line 7).

Usually these are some kind of 'auxiliary' codes that are created in cases where a link could not be found. We find that our linked data set covers 82 to 90 percent of persons in Business Register plants (see line 4). Although there are many plants in Business Register that are not linked with Employment Statistics (see line 7), they are usually quite small.

Plant-level data derived from Employment Statistics, PESF, can be linked further with other sources. We have linked PESF with Financial Statements Statistics (FSS) by using firm codes. If the firm is not covered by the Financial Statistics survey, information on profits is obtained from the registers compiled for taxation purposes. The linking has been done for the purpose of examining how the owner firm's profitability influences the job and worker flows. Table 6 illustrates how the sample changes due to linking of the PESF and FSS data. Because of concerns about the comparability of the profitability measures, we restricted our analysis to limited liability companies. Furthermore, for this analysis we excluded the finance and transportation sectors because of their poor coverage in the Financial Statements Statistics data.

**TABLE 6 HERE**

For the purpose of analyzing the relationship of the R&D intensity and job and worker flows we have linked PESF with R&D Statistics (RDS) information using firm codes. This investigation focuses on the manufacturing sector. Table 7 shows that the PESF data includes about 350 000 persons employed in manufacturing plants (about 15 000 plants). There are usually more than 200 000 persons in manufacturing firms covered by R&D Statistics (see line 2). The sub-sample, which is obtained by linking PESF and R&D Statistics data, accounts for some 70 percent of employment in the original PESF data. On the other hand, the R&D firms that are linked to at least one manufacturing plant have more employment than manufacturing R&D firms. The explanation for this outcome is that some R&D intensive service firms have also manufacturing plants. Thus line 4 in Table 7 includes also persons who work in a manufacturing plant that is owned by a service firm.

**TABLE 7 HERE**

Figure 2 shows the net employment change in the non-farm business sector in three data sets, full Employment Statistics data (ES), National Accounts data (NA), and our job and worker flow data set (PESF). The National Accounting figures are constructed from several sources, which may vary by sectors. The figure clearly shows that our data set tracks the changes in Employment Statistics data fairly well. Compared to the national accounts figures, both of them have slightly different timing in their cyclical variation. In Employment Statistics data and in our sample, the bottom of the recession was reached in 1992, when employment dropped by approximately 12 percent. In National Accounts data, the bottom was in the following year and the drop slightly smaller. The recovery started in Employment Statistics data in 1994, whereas in National Accounts it started in the following year. These differences are most likely caused by differences in the time that the statistics refer to. Employment Statistics records employment during the last week of the year, whereas the figure in the National Accounts refers to the annual average.

**FIGURE 2 HERE**

# 4. Illustrations of Research Uses of Linked Data

## 4.1 The Roles of Employer and Employee Characteristics for Plant Productivity

Our first example of research conducted using linked employer-employee data for Finland deals with productivity and wage profiles. Their relationship has been under much discussion in theoretical and econometric studies. Models of firm-specific human capital imply that in the early career wage exceeds productivity, but the productivity profile is steeper than the wage profile so that in the later career productivity exceeds the wage. If skills are not firm-specific but general, wage and productivity profiles should be similar. On the other hand, incentive wage models suggest that to keep working incentives high to the retirement age, wages should in the early career be below productivity and in the later career above productivity. In countries with strong labor unions, wages may also rise with seniority because of the bargaining power of the insiders. The compensation systems may guarantee steady wage increases that are not directly related to productivity. Another impact from personal characteristics to productivity comes through education. Skills acquired in education before the working career should be reflected in a productivity profile that starts at a higher level than without education. It is likely that skills acquired either through education or experience are complementary to the capital input and/or technology. With a newer capital stock, a given skill should give higher productivity. It is therefore necessary to control the age or vintage of the plant.

There are only a few studies where productivity and wage profiles have been estimated from a production function (e.g. Hellerstein, Neumark and Troske, 1999). We follow a related, but slightly different approach by analyzing total factor productivity directly. The estimated productivity and earnings equations have the following general specification

$$\ln(Y_{it}) = \alpha_0 + \alpha_i + \beta_1 \ln(AGE_{it}) + \beta_2 [\ln(AGE_{it})]^2 + \beta_3 \ln(SEN_{it}) + \beta_4 [\ln(SEN_{it})]^2 \\ + \beta_5 EDUY_{it} + \beta_6 [EDUY_{it})]^2 + X_{it}\delta + \varepsilon_{it}$$

In the productivity equation the dependent variable *Y* is the log of the multilateral total factor productivity indicator ln(TFP), and in the earnings equation the log of average hourly wage ln(WAGE). The labor characteristics variables are the log of the average age of the employees (AGE), the log of average seniority years (SEN), and the average years spent in schooling by the employees (EDUY) and squares of these variables. The only difference between productivity and wage specifications is that the square of education years was not significant in the TFP equations and was dropped from them, but it had a clearly significant coefficient in the wage models. These equations were estimated using plant panel data with OLS, and with random and fixed effects estimators to control for the plant specific fixed effects $\alpha_i$ .

In OLS and random effect estimations with levels the other plant-specific control variables $X_{it}$ included geographical location, the ratio of rents paid to the value of machinery, an indicator of foreign ownership, recent investments, ‘shadow of the death’, average hours per worker, and capacity utilization. In order to control the age of the establishment we use the plant generation variables (GENA-GENF). A linear trend was included in the models for the whole period and it was allowed to vary across 4-digit industries. In addition, the recession period 1991-1994 was indicated with a dummy variable, and dummies were included for the 4-digit industries. In the fixed effect model the time invariant variables, geographical location and industry dummies were dropped. We present here figures that are based on OLS estimation.

(More detailed results are presented in Ilmakunnas, Maliranta, and Vainiomäki (1999), where also the influence of worker turnover on productivity has been examined.)

Figure 3 shows how the age-productivity profile is altered when the other labor and plant characteristics are controlled. Curves (1)-(4) are alternative age-productivity profiles, and curve (5) is an age-wage profile that corresponds to productivity profile (4). The levels of the productivity and wage profiles should not be compared, since wage and total factor productivity are not measured in the same units. Instead, comparison of the slopes of the profiles can give an indication on which theories seem to be supported by the data.

**FIGURE 3 HERE**

Profile (1) shows the relationship between total factor productivity and age, when education, seniority and plant vintage are not included in the model. This profile reaches its peak at 33 years. One potential explanation for the success of the plants that have young personnel comes from the fact that generally the newer generations are more educated than the older ones. The years spent in education are controlled in profile (2). Some of the difference in the productivity performance between the plants where the average age of personnel is, say, 35 years and the plants where the average age is 50 years, can be accounted for by differences in education. It is worth noting that in terms of productivity the returns to schooling seem to be substantial, some 8 percent annually. In profile (3), the age of the plant is controlled with a dummy variable denoting the generation of the plant, and, furthermore, we have allowed for different trends for each generation. The relative performance of the plants that have older personnel improves noticeably. Finally, profile (4) demonstrates the relationship between age and productivity when log of seniority and its square are included in the model. Because long seniority years appear to affect productivity negatively and seniority years and age are positively correlated, the relative performance of the plants having personnel in advanced years improves after the control of seniority. Profile (4) seems to suggest that age in itself is not necessarily a burden in terms of low productivity, but rather the factors that are often associated with it: technology that is out of date, low turnover of workers, and low education. Age-wage profiles have fairly similar concave forms as the productivity profiles, which is consistent with the implications of general human capital. When all the factors are controlled, wage reaches its peak at the age of 40. This is shown in Figure 3 by wage profile (5).

Next, we examine the impact of seniority on productivity and wage. Figure 4 shows the productivity and wage profiles. Productivity profile (1) is from an estimation, where age and education are controlled, but plant vintage is excluded. Productivity increases initially fast with experience in the plant. However, it reaches its peak already at 2.5 years, and declines slowly over time thereafter. Long seniority appears to affect productivity negatively. The fact that skills are acquired fast seems to indicate that they are not firm-specific. The corresponding wage profile (2) is quite different: wages keep on increasing with seniority. The different forms of the wage and experience profiles can be interpreted in alternative ways. On one hand, the result supports the view that seniority-based wage may be used for keeping productivity incentives high. Another interpretation is that there is insider influence on wage determination, which is not related to productivity. When the plant generation variables are included (productivity profile (3)), the seniority-productivity profile shifts up. It peaks slightly later, at 3.8 years, and declines more slowly. The main conclusion remains, however, intact: productivity starts declining fairly early in the career, but wage (profile (4)) keeps on increasing. The coefficients of the plant generation variables showed that newer vintages have higher productivity. It seems that in profile (1) the seniority variables have picked up some of the vintage effect. Older plants have higher average seniority: in 1988 it was 3.3 years in generation A, in contrast to 11.1 years among generation F plants. When the plant generation is controlled, the productivity profile reflects the true influence of seniority better. Generally, the results indicate that although there are firm-specific skills, they are fairly fast adopted and they also erode fairly quickly.

**FIGURE 4 HERE**

## 4.2 Job and Worker Flows in the Finnish Business Sector

Our second example of applications of linked employer-employee data is an examination of job and worker flows. Key questions in this area of research have been the cyclicality of the flows and the differences between sectors. A ‘stylized fact’, which has been used as a basis of many macroeconomic theories, is the countercyclicality of job reallocation. This result is mainly based on information on US manufacturing (Davis, Haltiwanger and Schuh, 1996). Another interesting issue is the relationship between job and worker flows. Their examination with the same underlying data has been possible only in some special cases. Our data are fairly similar to those used in the study of job and worker flows in the other Nordic countries; see, e.g., Albæk and Sørensen (1998) and Bingley, Eriksson, Westergård-Nielsen, and Werwatz (1999) for Denmark, Persson (1999) for Sweden, and Barth and Dale-Olsen (1997) and Salvanes (1999) for Norway.

We present here results on the Finnish non-farm business sector in the period 1988-1996. Figures 5 and 6 show the flow rates for the whole sector. Corresponding figures for the main industries (Ilmakunnas and Maliranta, 2000b, present results for the following industries: manufacturing (including mining and energy), wholesale and retail trade, hotels and restaurants, finance, construction, and business services) show fairly similar cyclical variations, although the levels of the flows vary by industry. The flows are higher in the service industries and in construction than in manufacturing and finance. The main implications that can be drawn are the following. Job creation is procyclical and job destruction countercyclical, but they vary so symmetrically that the job reallocation rate is more or less acyclical (the correlation of JR and NET is 0.21). This result is obtained in most of the industries, too. Only in hotels and restaurants is JR negatively correlated with NET. In some cases (finance, business services), there is actually fairly strong procyclicality of job reallocation. In contrast to this, the worker inflow rate (or hiring rate) WIF and outflow rate (or separation rate) WOF vary so that the procyclicality of the inflow rate is much stronger than the countercyclicality of the outflow rate. As a result, the worker turnover rate WF is strongly procyclical. This means that firms adjust to cyclical changes more by adjusting their inflow rate than by adjusting the outflow rate. Finally, it can be noted that the churning rate varies procyclically, i.e. in the downturn, excessive worker turnover is decreased. The same also happens to excess job reallocation.

**FIGURE 5 HERE**
**FIGURE 6 HERE**

Plants were divided into groups according to various background characteristics (see Ilmakunnas and Maliranta, 2000a,b). We briefly summarize here some results without presenting them in tables or figures. The flows vary by plant age and size so that younger and smaller plants have higher job and worker turnover rates. However, churning is high in larger plants. As to worker characteristics, highest worker and job turnover happens in plants with highly educated employees or with employees who have a low education level. On the other hand, high churning is related to a high educational level. The turnover rates are highest for plants that have low wage, low capital intensity, or low export intensity.

In some analyses the plant-level flow data has been linked to firm-level data. Table 8 shows how the job and worker flows are associated with the profitability of the firm. This analysis is made by linking the PESF data with Financial Statistics using firm codes. The plants were classified by profitability into 5 groups, which all cover 20 percent of the employment in the plants. (Employment share is indicated by column W and the employment share weighted net employment change by column WNET in Tables 8 and 9.) A high profitability of the firm is closely and positively associated with net job creation. This result derives very much from the fact that the job destruction rate is particularly low among high profitability firms. The worker and job turnover rates as well as the excessive job reallocation and churning rates are also relatively low among them. Furthermore, we notice that worker outflow to unemployment (WOFU) is relatively low as well. On the other hand, it seems that high profitability firms absorb unemployment at a lower rate than the less profitable ones; WIFU is low in this group.

Similar type of analysis has been made in another study for the purpose of studying the role of R&D intensity of the firm in the job and worker flows (Maliranta, 2000). Plants in the PESF data have been linked with the R&D Statistics survey with firm codes. By this procedure we are able to classify plants into groups according to the R&D intensity (R&D expenditures per sales) of the owner firm. Table 9 indicates that the R&D intensity is positively correlated with the net and gross job creation rates at the micro level. We notice that, for example, the excessive job reallocation and churning rates are highest among high R&D and low R&D intensity plants. High R&D intensive plants appear to absorb unemployed at a higher rate than low R&D intensive ones.

**TABLE 8 HERE**

**TABLE 9 HERE**

## 5. Conclusions

We have described the development of linked employer-employee data for Finland and their application in labor market analysis. In countries like Finland where register-based information on the whole population of employees and plants is available, it is feasible to create linked data by combining information from different sources and thereby obtain a better picture of the labor market. Even with this kind of data sets, the matching of the data is not an easy task. Different practices in the various statistics, and data needs in the research lead to incompleteness in the linking and to loss of data. The final data sets are therefore non-random samples of the original data, but the properties of the sample are unknown. This property of the data has to be taken into account when interpreting the results, although its formal analysis is difficult. On the other hand, our examples illustrate that the data sets have a good coverage of the whole population of plants and employees.

For other research topics one would perhaps end up with different kinds of data sets. It seems that it is difficult to form a processed data set, which could be used for multiple purposes. Instead, the best policy might be to form a good infrastructure for making different kinds of linkings from the raw register data. Developments in computer technology have made it possible to process large quantities of micro data to different kinds of data sets at a reasonable cost.

## References

Albæk, K. and Sørensen, B.E., “Worker flows and job flows in Danish manufacturing, 1980-91”, *Economic Journal* 108, 1998, 1750-1771

Barth, E. and Dale-Olsen, H., *Jobb- og arbeidskraftsmobilitet i Norge,* Institute for Social Research, Oslo, Rapport 97:17, 1997

Bingley, P., Eriksson, T., Werwatz, A. and Westergård-Nielsen, N., “Beyond ’manucentrism’ – some fresh facts about job and worker flows”, *CLS Working Paper* 99-09, 1999

Burgess, S., Lane, J., and Stevens, D., “Job Flows, Worker Flows and Churning”, *Journal of Labor Economics* 18, 2000, 473-502

Caves, D. W., Christensen, L. R., and Trethway, M. W., “U.S. Trunk Air Carriers, 1972-1977: A Multilateral Comparison of Total Factor Productivity Growth”, in T. G. Cowing and R. E. Stevens, eds., *Productivity Measurement in Regulated Industries*, New York: Academic Press, 1981

Caves, D. W., Christensen, L. R., and Diewert, W. E., “Multilateral Comparisons of Output, Input and Productivity Using Superlative Index Numbers”, *Economic Journal* 92, 1982, 73-86

Davis, S. J., Haltiwanger, J. C., and Schuh, S., *Job Creation and Destruction*, Cambridge: MIT Press, 1996

Griliches, Z. and Regev, H., “Firm Productivity in Israeli Industry 1979-1988”, *Journal of Econometrics* 65, 1995, 175-203

Haltiwanger, J., Lane J., Spletzer J., Theeuwes J. and Troske K., eds., *The Creation and Analysis of Employer-Employee Matched Data*, Amsterdam: Elsevier Science, 1999

Hellerstein, J. K., Neumark, D., and Troske, K. R., “Wages, Productivity, and Worker Characteristics: Evidence from Plant-level Production Functions and Wage Equations”, *Journal of Labor Economics* 17, 1999, 409-446

Ilmakunnas, P. and Maliranta, M., *Työpaikkojen syntyminen ja häviäminen ja työvoiman vaihtuvuus* (*Job Creation and Destruction and Worker Turnover*), Helsinki: Ministry of Labour, Studies in Labour Policy, No. 209, 2000a

Ilmakunnas, P. and Maliranta, M., *The Turnover of Jobs and Workers in Finland*, manuscript, 2000b

Ilmakunnas, P., Maliranta, M., and Vainiomäki, J., “The roles of employer and employee characteristics for plant productivity”, *Helsinki School of Economics and Business Administration*, *Working Papers* W-223, 1999

Maliranta, M., “Plant Productivity in Finnish Manufacturing — Characteristics of High Productivity Plants”, *Research Institute of the Finnish Economy, Discussion Papers* No. 612, 1997

Maliranta, M, “Factors of Performance by Plant Generations. Some Findings from Finland”, in Biffignandi, S., ed., *Micro- and Macrodata on Firm: Statistical Analysis and International Comparisons*, Heidelberg: Physica-Verlag, 1999, 391- 424

Maliranta, M. ‘Job creation by supporting technology advances; Evidence from Finnish plants’, in Asplund, Rita, ed., *Public R&D Funding, Technological Competitiveness, Productivity, and Job Creation*, ETLA B 168, 2000, 87-128

Persson, H., “Job flows and worker flows in Sweden 1986-1995”, in H. Persson: *Essays on Labour Demand and Career Mobility*, Swedish Institute for Social Research, Dissertation Series 40, 1999

Salvanes, K.G., “Employment policies at the plant level: Job and worker flows for heterogeneous labour and heterogeneous plants in Norway”, *Norwegian School of Economics and Business Administration, Discussion paper* 6/99, 1999

Vainiomäki, J, “Technology and Skill Upgrading: Results from Linked Worker-Plant Data for Finnish Manufacturing”, in Haltiwanger, J. et. al., eds., 1999, 115-145

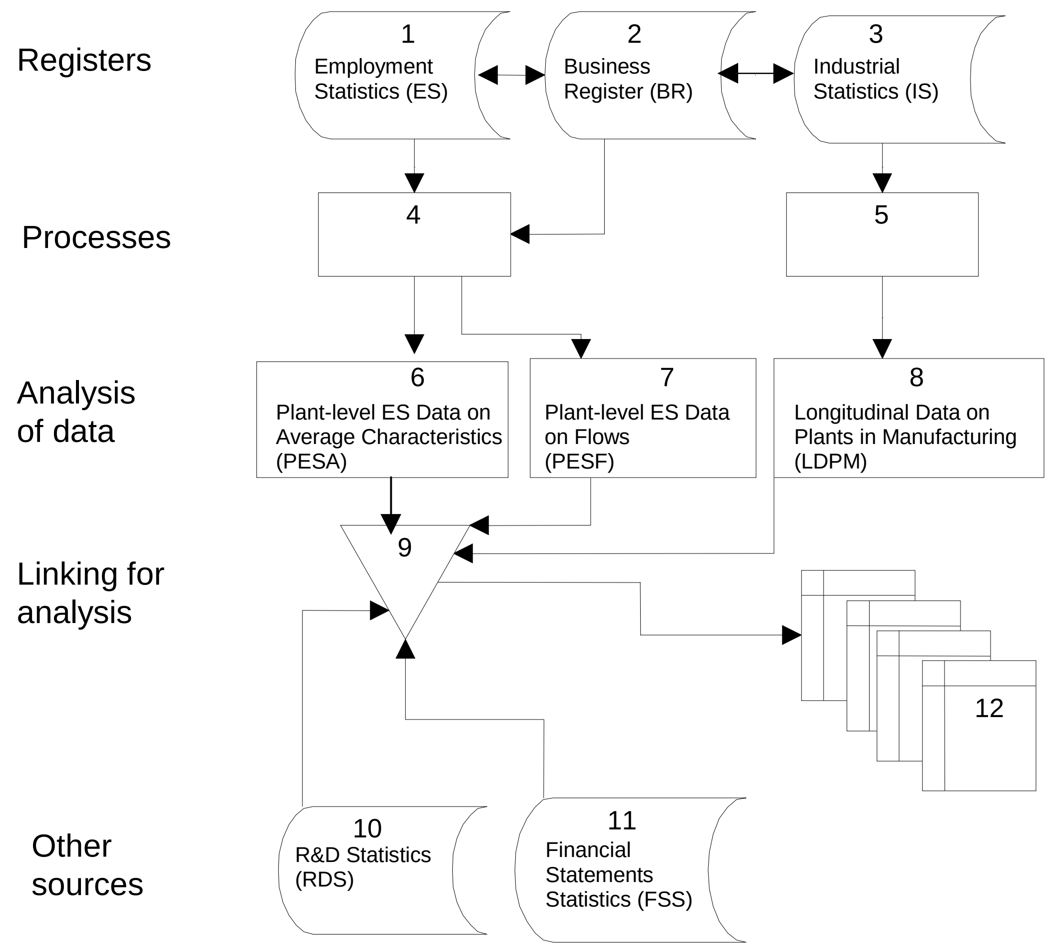
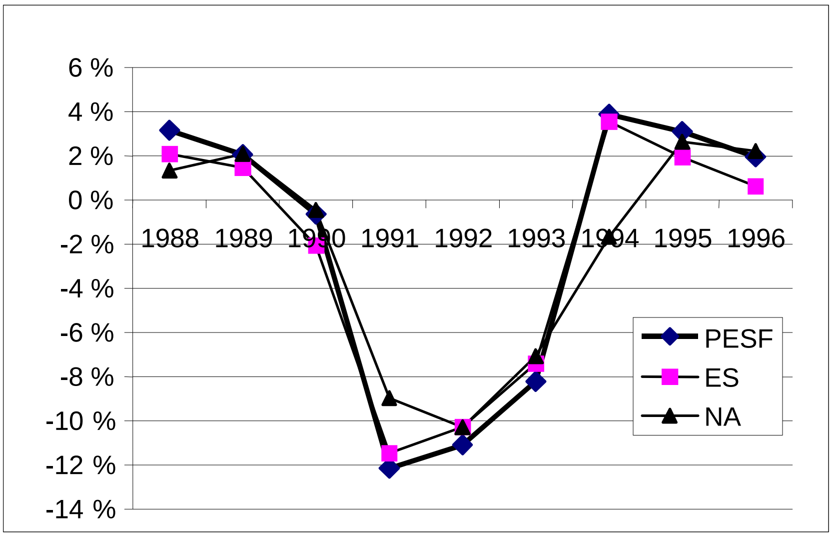



**Figure 1: Registers and the linking process**




Note: PESF: Plant-level ES data on flows, ES: Employment Statistics, NA: National Accounts

**Figure 2: Net employment change in different statistics, business sector**

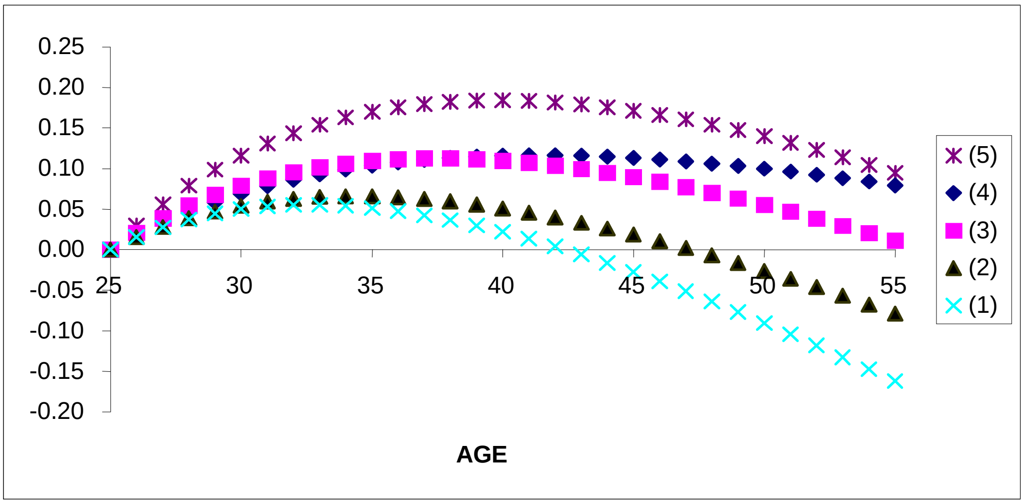



**Figure 3: Productivity and wage profiles according to average age**

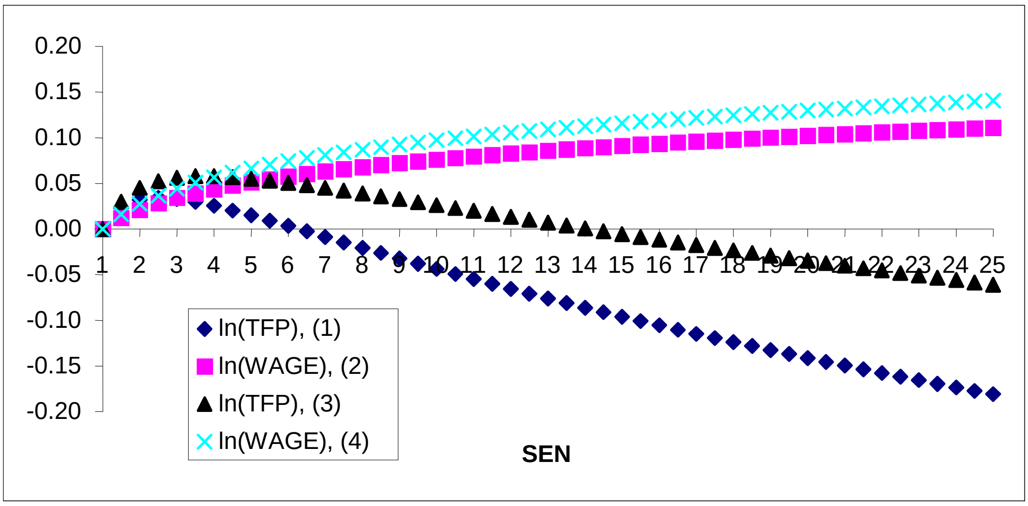
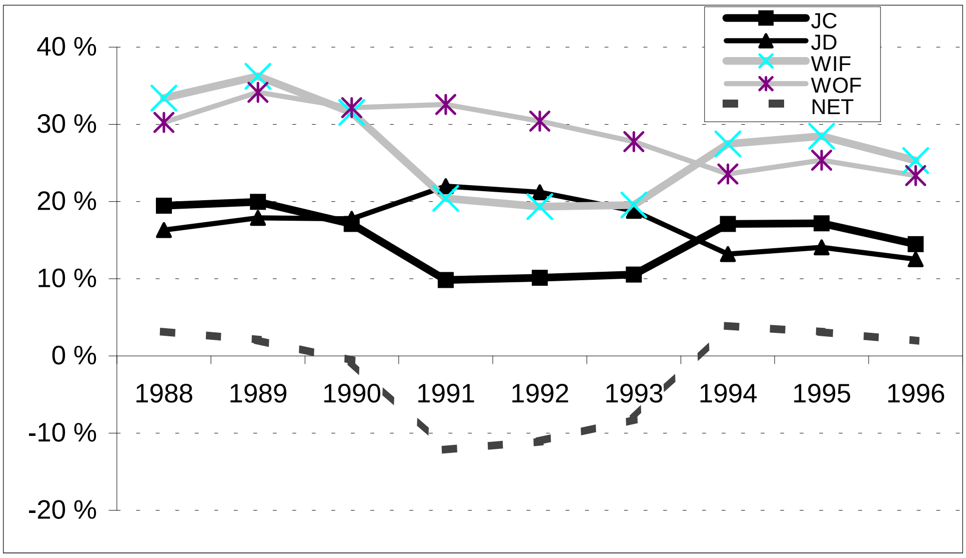
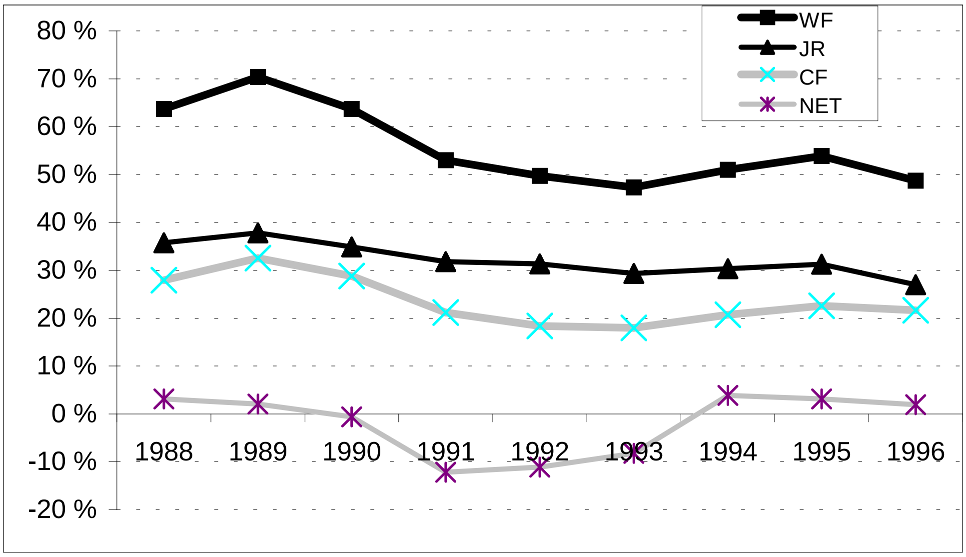



**Figure 4: Productivity and wage profiles according to average seniority**




Note: JC: job creation, JD: job destruction, WIF: worker inflow, WOF: worker outflow, NET: net employment change

**Figure 5: Gross job and worker flow rates in the business sector**




Note: WF: worker flow (turnover), JR: job reallocation (turnover), CF: churning flow, NET: net employment change

**Figure 6: Job and worker turnover rates in the business sector**

**Table 1. Main Features of Different Registers**

| Register | Units of observation | Coverage | Information content |
|---|---|---|---|
| Employment Statistics (ES) | Individuals | Whole working age population | Age, sex, education, seniority, earnings, marital status, area of residence, home ownership, etc.; employer in the last week of the year |
| Business Register (BR) | Firms and plants | All firms and employers subject to VAT, some minimum size restrictions (e.g. at least 0.5 employees); all industries | Employment, wage bill, sales, area |
| Industrial Statistics (IS) | Plants | Plants with at least 5 employees (up to 1994; from 1995 smaller sample); manufacturing, mining and energy | Gross output, value added, purchase of inputs, investment, employment, hours worked, ownership, area, etc. |
| Financial Statements Statistics (FSS) | Firms | Firms with at least 20 employees always included; information on smaller firms from tax registers; all industries | Balance sheet, income statement |
| R&D Statistics (RDS) | Firms, (groups of enterprises) | Firms with at least 100 employees always included, some smaller firms; manufacturing and some service industries | R&D expenditure, sales |

**Table 2: Matching Workers to Plants: Employment Statistics and Industrial Statistics.**

| | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 |
|---|---|---|---|---|---|---|---|
| 1. All persons in ES | 446125 | 445986 | 442819 | 387695 | 352048 | 346680 | 369583 |
| 2. ES persons in IS plants | 353922 | 354629 | 352714 | 318476 | 287393 | 273224 | 285388 |
| * share of line 1 | 79.3 % | 79.5 % | 79.7 % | 82.1 % | 81.6 % | 78.8 % | 77.2 % |
| 3. Linked persons | 283831 | 303100 | 311883 | 279383 | 257425 | 247808 | 255928 |
| * share of line 4 | 65.0 % | 70.8 % | 75.6 % | 72.9 % | 74.7 % | 76.3 % | 78.5 % |
| 4. Persons in IS | 436484 | 427950 | 412737 | 383428 | 344388 | 324765 | 326217 |
| 5. Plants in IS | 6316 | 6237 | 6101 | 6480 | 5941 | 5595 | 5379 |
| 6. Linked plants | 5530 | 5651 | 5565 | 5831 | 5243 | 4943 | 4821 |
| * share of line 5 | 87.6 % | 90.6 % | 91.2 % | 90.0 % | 88.3 % | 88.3 % | 89.6 % |

Notes for the rows:

1. All persons from Employment Statistics database: aged at least 15, industry of employment is manufacturing, plant code not empty.

2. Persons whose plant codes in ES and IS are the same, ES codes compared to the list of plant codes of active production plants in IS.

3. Persons fulfilling restrictions to be included in calculations for plant-level variables: a) employed, b) wage and salary earner, c) length of employment > 1 month, d) monthly wage available and between min-max bounds.

4. All persons in active production plants in IS.

5. Active production plants in IS.

6. Plants that had at least one linked worker fulfilling the restrictions required on line 3.

**Table 3: Descriptive Statistics for plants at various data steps.**

| Data set | Year | Number of persons | Share of data set A | Number of plants | Share of data set A | Average size | Nominal value added per hour | Wage per employee |
|---|---|---|---|---|---|---|---|---|
| A | 1990 | 434391 | 100 % | 7182 | 100 % | 60 | 169 | 68 |
| | 1994 | 344756 | 100 % | 6601 | 100 % | 52 | 235 | 80 |
| B | 1990 | 412737 | 95.0 % | 6101 | 84.9 % | 68 | 178 | 67 |
| | 1994 | 326217 | 94.6 % | 5379 | 81.5 % | 61 | 248 | 78 |
| C | 1990 | 370320 | 85.3 % | 5005 | 69.7 % | 74 | 183 | 67 |
| | 1994 | 296543 | 86.0 % | 4317 | 65.4 % | 69 | 261 | 79 |
| D | 1990 | 347387 | 80.0 % | 4536 | 63.2 % | 77 | 183 | 67 |
| | 1994 | 279181 | 81.0 % | 3882 | 58.8 % | 72 | 261 | 79 |

**Table 4: Descriptive statistics on worker characteristics**

| | | Plants with worker characteristics estimate | | | | | Plants in regression analysis (data set D) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | YEAR | Number. of plants | MEAN | MED | P95 | P5 | Number. of plants | MEAN | MED | P95 | P5 |
| AGE | | | | | | | | | | | |
| | 1990 | 5466 | 37.8 | 37.8 | 45.3 | 30.2 | 4536 | 37.9 | 38.0 | 45.2 | 30.4 |
| | 1994 | 4755 | 39.1 | 39.3 | 46.2 | 31.7 | 3882 | 39.2 | 39.4 | 46.1 | 31.9 |
| SEN | | | | | | | | | | | |
| | 1990 | 5466 | 8.3 | 7.7 | 17.1 | 1.9 | 4536 | 8.5 | 8.0 | 17.3 | 2.0 |
| | 1994 | 4755 | 9.4 | 9.0 | 18.3 | 1.8 | 3882 | 9.7 | 9.2 | 18.4 | 1.9 |
| EDUY | | | | | | | | | | | |
| | 1990 | 5466 | 10.6 | 10.5 | 11.9 | 9.6 | 4536 | 10.5 | 10.5 | 11.8 | 9.6 |
| | 1994 | 4755 | 10.8 | 10.7 | 12.4 | 9.7 | 3882 | 10.8 | 10.7 | 12.3 | 9.8 |

**Table 5: Matching of workers and plants in the analysis of flows, business sector, linking ES and BR**

| **Persons, 000s** | **1988** | **1990** | **1992** | **1994** | **1996** | **1998** |
|---|---|---|---|---|---|---|
| 1. All persons in ES | 1462 | 1454 | 1174 | 1124 | 1150 | |
| 2. All persons in BR | 1289 | 1349 | 1103 | 976 | 1044 | 1142 |
| * share of line 1. | 88 % | 93 % | 94 % | 87 % | 91 % | |
| Persons in linked plants | | | | | | |
| 3. Persons in ES | 1138 | 1161 | 908 | 875 | 929 | |
| * share of line 1. | 78 % | 80 % | 77 % | 78 % | 81 % | |
| 4. Persons in BR | 1124 | 1101 | 904 | 820 | 924 | |
| * share of line 2. | 90 % | 86 % | 85 % | 82 % | 86 % | |
| **Plants, 000s** | **1988** | **1990** | **1992** | **1994** | **1996** | **1998** |
| 5. Plants in ES | 98.5 | 106.3 | 92.4 | 92.3 | 96.2 | |
| * share of line 6. | 72 % | 71 % | 65 % | 55 % | 50 % | |
| 6. Plants in BR | 137.4 | 150.1 | 142.5 | 168.4 | 193.5 | 204.7 |
| Linked plants, 000s | | | | | | |
| 7. Plants in ES & BR | 92.70 | 97.73 | 83.82 | 87.34 | 91.86 | |
| * share of line 5. | 94 % | 92 % | 91 % | 95 % | 96 % | |
| * share of line 6. | 67 % | 66 % | 59 % | 51 % | 47 % | |

**Table 6: Matching of workers and plants in the analysis of flows, business sector, linking PESF and FSS**

| **Persons, 000s** | **1994** | **1995** | **1996** |
|---|---|---|---|
| 1. Persons in PESF | 721 | 748 | 774 |
| 2. Persons in FSS | 646 | 728 | 778 |
| * share of line 1. | 90 % | 97 % | 100 % |
| Persons in linked plants | | | |
| 3. Persons in PESF | 520.0 | 608.5 | 654.3 |
| * share of line 1. | 72 % | 81 % | 84 % |
| 4. Persons in FSS | 587 | 637 | 681 |
| * share of line 2. | 91 % | 88 % | 88 % |

Note: Finance and transportation sectors are excluded. FSS is firm-level data.

**Table 7: Matching of workers and plants in the analysis of flows, manufacturing sector, linking PESF and RDS**

| **Persons, 000s** | **1994** | **1995** | **1996** |
|---|---|---|---|
| 1. Persons in PESF | 348 | 356 | 362 |
| 2. Persons in RDS | 232 | 258 | 181 |
| * share of line 1. | 67 % | 72 % | 50 % |
| Persons in linked plants | | | |
| 3. Persons in PESF | 246 | 257 | 255 |
| * share of line 1. | 71 % | 72 % | 70 % |
| 4. Persons in RDS | 263 | 283 | 205 |
| * share of line 2. | 113 % | 110 % | 113 % |

Note: Row 2 includes only persons in plants of manufacturing firms in RDS.
Row 4 includes persons in all firms in RDS that have at least one plant in PESF.
This comprises personnel in service firms in case they have manufacturing plants appearing in PESF.
RDS is firm-level data.

**Table 8: Flow rates (%) by profitability in the business sector (finance and transportation excluded), 1994-96**

| Net profit-% | Worker inflow WIF | Worker outflow WOF | Worker flow WF = WIF+WOF | Job creation JC | Job destruc-tion JD | Job reallo-cation JR = JC+JD | Churning flow CF = WF-JR | Excess job reallo-cation EJR = JR-\|NET\| | Net change NET = JC-JD = WIF-WOF | Employ-ment share weighted net change WNET | Worker inflow from unem-ployment WIFU | Worker outflow to unem-ployment WOFU | Net flow from unem-ployment UNET = WIFU-WOFU | Employ-ment share W |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| High | 25.0 | 17.3 | 42.3 | 14.7 | 7.1 | 21.8 | 20.4 | 14.2 | 7.6 | 1.5 | 4.4 | 2.7 | 1.7 | 20 |
| 2 | 25.1 | 18.6 | 43.6 | 14.9 | 8.4 | 23.2 | 20.4 | 16.7 | 6.5 | 1.3 | 5.0 | 3.6 | 1.4 | 20 |
| Medium | 25.0 | 20.1 | 45.1 | 14.0 | 9.1 | 23.1 | 22.0 | 18.2 | 4.9 | 1.0 | 4.8 | 4.3 | 0.5 | 20 |
| 4 | 26.8 | 23.1 | 49.9 | 14.9 | 11.1 | 26.0 | 23.9 | 22.3 | 3.7 | 0.7 | 5.8 | 5.5 | 0.3 | 20 |
| Low | 27.6 | 27.9 | 55.5 | 15.9 | 16.2 | 32.0 | 23.4 | 27.4 | -0.3 | -0.1 | 6.8 | 7.6 | -0.8 | 20 |

Note: Figures are averages of annual figures over the period 1994-96.

**Table 9: Flow rates (%) by R&D intensity in manufacturing, 1994-96**

| R&D intensity | Worker inflow WIF | Worker outflow WOF | Worker flow WF = WIF+WOF | Job creation JC | Job destruc-tion JD | Job reallo-cation JR = JC+JD | Churning flow CF = WF-JR | Excess job reallo-cation EJR = JR-\|NET\| | Net change NET = JC-JD = WIF-WOF | Employ-ment share weighted net change WNET | Worker inflow from unem-ployment WIFU | Worker outflow to unem-ployment WOFU | Net flow from unem-ployment UNET = WIFU-WOFU | Employ-ment share W |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 % – 0.5 % | 18.9 | 18.6 | 37.5 | 9.3 | 9.0 | 18.3 | 19.2 | 16.8 | 0.3 | 0.1 | 2.9 | 4.0 | -1.1 | 27.0 |
| 0.5 % – 1 % | 16.9 | 18.9 | 35.8 | 8.1 | 10.1 | 18.2 | 17.6 | 16.1 | -2.0 | -0.5 | 2.2 | 3.5 | -1.3 | 27.9 |
| 1 % – 2 % | 17.5 | 11.8 | 29.2 | 10.3 | 4.6 | 15.0 | 14.3 | 9.3 | 5.7 | 0.8 | 3.8 | 2.6 | 1.2 | 15.3 |
| 2 % – 5 % | 17.1 | 14.0 | 31.1 | 9.8 | 6.8 | 16.6 | 14.5 | 13.6 | 3.0 | 0.5 | 3.4 | 2.2 | 1.1 | 15.1 |
| 5 % – 50 % | 32.7 | 22.2 | 54.9 | 18.6 | 8.2 | 26.7 | 28.2 | 16.3 | 10.4 | 1.5 | 4.0 | 1.5 | 2.5 | 14.7 |

Note: Figures are averages of annual figures over the period 1994-96.